Light Fields for Face Analysis

Chiara Galdi, Valeria Chiesa, Christoph Busch, Paulo Lobato Correia, Jean-Luc Dugelay, Christine Guillemot

sensors

MDPI

Article

Light Fields for Face Analysis

Chiara Galdi [1,*], Valeria Chiesa [1,*], Christoph Busch [2], Paulo Lobato Correia [3], Jean-Luc Dugelay [1] and Christine Guillemot [4]

1 Department of Digital Security, EURECOM, 06560 Sophia Antipolis, France; jean-luc.dugelay@eurecom.fr
2 Hochschule Darmstadt, 64295 Darmstadt, Germany; Norwegian University of Science and Technology, 2815 Gjøvik, Norway; christoph.busch@ntnu.no
3 Instituto de Telecomunicacoes/Instituto Superior Tecnico, Universidade de Lisboa, 1049 - 001 Lisboa, Portugal; plc@lx.it.pt
4 Institut National de Recherche en Informatique et en Automatique, 35042 Rennes, France; christine.guillemot@inria.fr
* Correspondence: galdi@eurecom.fr (C.G.); chiesa@eurecom.fr (V.C.)

Received: 30 April 2019; Accepted: 12 June 2019; Published: date

Abstract: The term "plenoptic" comes from the Latin words plenus ("full") + optic. The plenoptic function is the 7-dimensional function representing the intensity of the light observed from every position and direction in 3-dimensional space. Thanks to the plenoptic function it is thus possible to define the direction of every ray in the light-field vector function. Imaging systems are rapidly evolving with the emergence of light-field-capturing devices. Consequently, existing image-processing techniques need to be revisited to match the richer information provided. This article explores the use of light fields for face analysis. This field of research is very recent but already includes several works reporting promising results. Such works deal with the main steps of face analysis and include but are not limited to: face recognition; face presentation attack detection; facial soft-biometrics classification; and facial landmark detection. This article aims to review the state of the art on light fields for face analysis, identifying future challenges and possible applications.

Keywords: survey; light field; face analysis

1. Introduction

This survey aims not only to collect and review studies showing how face analysis techniques can be adapted to the new light-field paradigm, but also to discuss the advantages of the use of light fields for face analysis compared to classical 2D imaging. The surveyed works are selected according to the following criteria: (i) the work deals with the use of light fields for face analysis; (ii) the work is published in peer-reviewed journal or conference proceedings.

Articles were retrieved thanks to Elsevier's abstract and citation database, namely Scopus. The advanced search tool allows identification of articles based on keywords and performance of analysis on the retrieved data. The keywords used in this case included: "light field"; "Lytro"; "Raytrix"; "plenoptic"; "face recognition/detection/landmark/liveness/presentation attack detection". No time range limits were imposed. The obtained list of papers was manually checked to exclude false matches (6 papers). Along with the articles retrieved from Scopus, all co-authors of this article contributed in adding any paper of interest not already included in the provided list.

Figure 1 illustrates the number of published papers by year, in the period 2002–2018, about light fields for face analysis.

The graph shows that first attempts of introducing the concept of "light fields" for face analysis were made in 2002 [1,2] and 2004 [3,4]. The authors of these works used classical 2D RGB images to create a light-field-inspired object to address face pose variation. The first work using images captured

by an actual plenoptic camera and thus using actual light-field face images was only published in 2013 [5].

The remainder of the survey is organized as follows: Section 2 introduces the light-field function and the novel features provided by the images captured by the devices implementing such technology, and the publicly available light-field face databases; Sections 3–5, review articles addressing the use of light fields for face landmark detection, face recognition, and face presentation attack detection, respectively. Section 6 discusses the main findings and implications for future research.

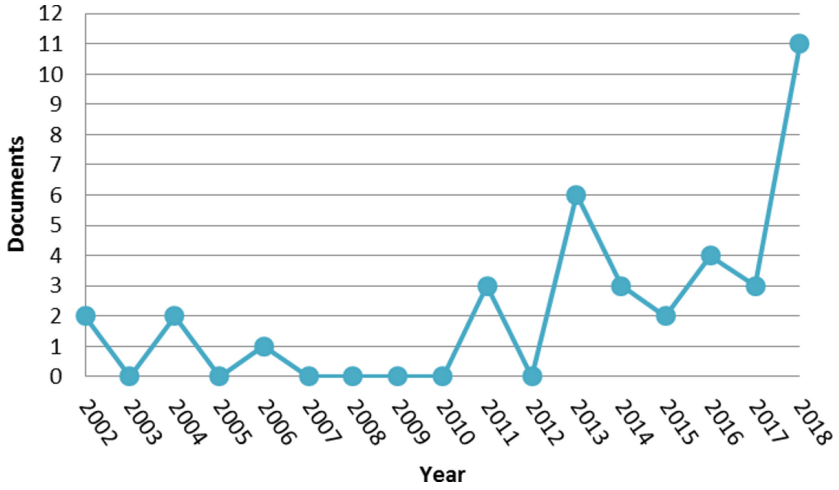


Figure 1. Graph of scientific papers published per year in the period 2002–2018 matching keywords for light fields and face analysis. From Scopus (https://www.scopus.com).

2. Background

2.1. Light Fields

Light-field imaging has been recently gaining in popularity due to its interest for a variety of computer vision applications, such as 3D modelling, object detection, classification, or recognition, with applications in computational photography, augmented reality, light-field microscopy, medical imaging, 3D robotics, and also biometric recognition. Light fields represent the radiance of light rays emitted by any point in a 3D scene along different orientations that is generally described by a 7-parameters function: $L(x, y, z, \theta, \phi, \lambda, t)$, where (x, y, z) are the 3D coordinates of the reference point, (θ, ϕ) are its angular coordinates, the point is observed at a particular wavelength λ and at a particular time t. Most of the plenoptic devices can collect only still images in visible spectrum, thus the parameters λ and t are fixed and often omitted. In some works ([6,7]), the authors adopt a different notation: the radiance is represented by a function $L(x, y, u, v)$ of 4 parameters, where (u, v) denote the angular or view coordinates (corresponding to different orientations of the light rays), and where (x, y) denote the spatial or pixel coordinates. A light-field image can therefore be seen as capturing an array of viewpoints (called sub-aperture images, see Figure 2) of the scene with varying angular coordinates u and v.

Camera rigs have been naturally constructed to capture the set of views, offering a high spatial resolution for each view but a low angular resolution (i.e., large baseline) [8]. Targeted applications include long range depth estimation, and augmented or virtual reality with immersive content. Single cameras mounted on moving gantries capturing the scene at regular time intervals have also been considered [9]. While camera rigs can be quite bulky and not easy to use, moving gantries are limited to capturing light fields of static scenes. Plenoptic cameras have also emerged based on novel optical designs [10,11]. Plenoptic cameras, thanks to an array of microlenses placed in front of the sensor, can be seen as multiplexing multiple low-resolution views in one 2D image sensor [11,12]. This is an efficient and easy way of capturing multiple viewpoints, even if the angular resolution is

achieved at the expense of a decreased spatial resolution, when compared with classical 2D cameras. Despite the small baseline they offer (small disparities between views), they turn out to be low-cost and easy-to-use devices for light-field captures, and therefore for 3D scene reconstruction. Figure 3 illustrates some examples of the devices described above.

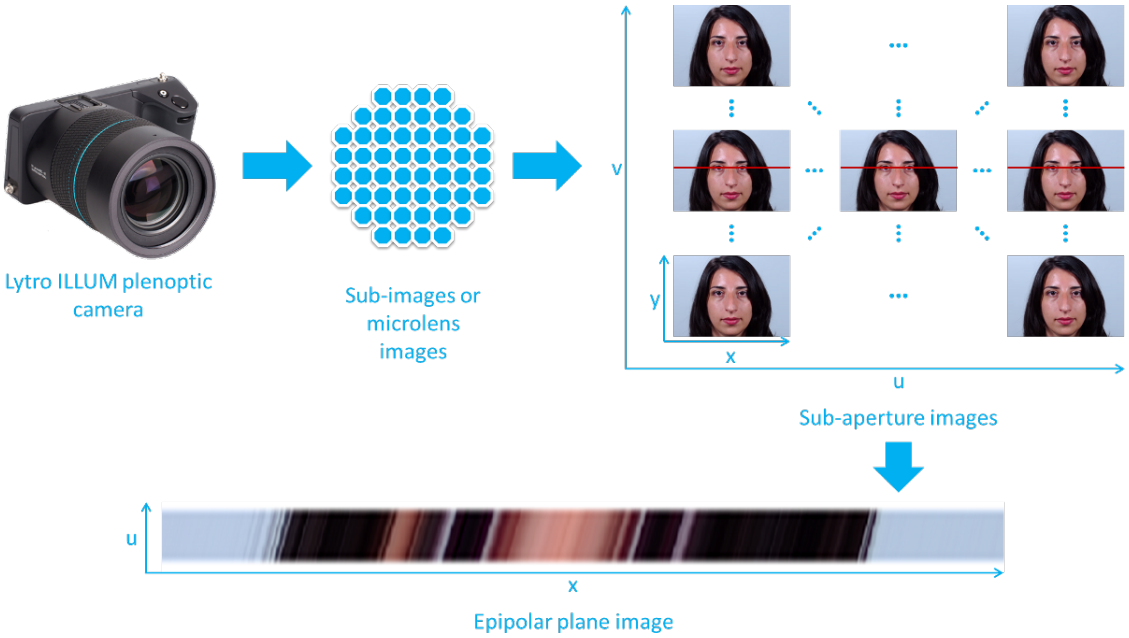
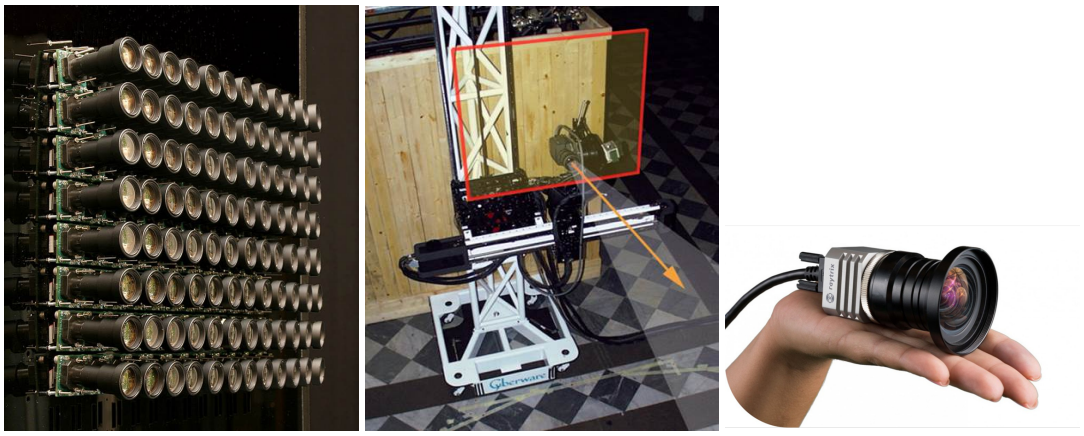


Figure 2. Epipolar plane image: process to obtain an epipolar plane image.



(**a**) (**b**) (**c**)

Figure 3. Examples of different implementations of light-field capturing devices: (**a**) camera rig mounting an array of eight by twelve cameras; (**b**) moving gantry shifting the camera in the area indicated by the red square; (**c**) plenoptic camera by Raytrix (image from https://raytrix.de/)

2.1.1. Epipolar Plane Images

The concept of epipolar plane image (EPI) has been firstly introduced by Bolles et al. in 1987 [13], for building a three-dimensional description of a static scene from a dense sequence of images. The idea is derived from the epipolar constraint in stereo vision.

An EPI represents a spatio-angular 2D slice of the 4D light field, cut through a horizontal or vertical stack of light-field views (e.g., a (x, u) slice corresponding to the horizontal red line in Figure 2). It is obtained by fixing one of the spatial coordinates (e.g., by fixing the spatial coordinate y in Figure 2) and one of the angular coordinates (v in Figure 2). The EPI shown in Figure 2 therefore, gives an

observation of a point at varying sub-apertures of the main lens or view positions u and at a given pixel location x.

Each observed 3D point, when projected in the EPI representation, traces lines with a certain slope that depends on the distance of the subject to the camera. The closer the camera is to the observed point, the steeper the slope is. This feature, as explained in the following paragraph, allows scene depth estimation.

2.1.2. Scene Depth from Light Fields

The rich scene description provided by light fields enables 3D scene geometry estimation and 3D scene reconstruction. Scene depth estimation methods from light fields can be broadly classified in two categories, depending on the light-field baseline (large or small disparities between views). A first category of methods follows the principles of stereo matching using robust patch-based block matching [14–16]. A second category analyses specific linear structures in EPIs [17,18] for depth computation from dense light fields. In fact, the pixels corresponding to the same 3D point in different views of a light-field image form a line in the EPI whose slope is proportional to the disparity value [**?**]. Hereafter, those lines will be referred to as level lines. Please note that while stereo methods allow estimating larger disparities, EPI-based methods are only suitable for densely sampled light fields. This is the case for the light-field images treated by the works reviewed in this article, since they have been mostly captured by Lytro ILLUM plenoptic cameras.

2.1.3. Refocusing from Light Fields

Densely sampled light fields make it possible to render images with shallow depth of field while controlling the focus depth. This set of images with different focusing depths is called the focal stack. Refocusing consists of defining a new light-field image $L'(x, y, u, v) = L(x - us, y - vs, u, v)$, where s is the refocus parameter defined in such a way that the regions of disparity $d = s$ in the light field L have zero disparity in the light field L'. A refocused image I^s with parameter s is computed by integrating the light rays over the angular dimension as [11]:

$$I^s(x, y) = \int\int_{\mathbb{R}} L(x - us, y - vs, u, v)\psi(u, v)\mathrm{d}u\mathrm{d}v. \quad (1)$$

where $\psi(u, v)$ represents the aperture of the imaging system, equal to 1 in the case of a full aperture. Therefore, refocusing is conceptually just a summation of shifted versions of the sub-aperture images over the entire $\psi(u, v)$ aperture.

2.2. Face Analysis

Light-field images, acquired with a plenoptic camera or with an array of conventional 2D cameras, include information about different viewpoints of the scene. This allows developing face analysis algorithms that exploit the multiple viewpoints to achieve improved performance. These algorithms can explore the light-field characteristics, including functionalities such as:

- *A posteriori* refocusing – As mentioned in Section 2.1, it is possible to refocus a posteriori the captured image at a given depth plane, thus rendering a 2D image where the objects at the selected plane appear in focus. This functionality is not available when using a conventional 2D camera and it can be very useful for face analysis algorithms, allowing improvement of the analysis of a previously out-of-focus region of interest.
- Disparity exploitation – Given a captured light-field image, it is possible to render a set of 2D images corresponding to a set of specified viewpoints. This provides disparity information, with the differences between corresponding points/face components in different viewpoints providing valuable information for face analysis.
- Depth exploitation—As mentioned in Section 2.1, it is possible to estimate depth information from the light-field image. Knowing the distance between face components and the camera,

it provides information about the scene geometry, useful for face analysis. An obvious usage of depth information is for presentation attack detection, where depth information can be used to determine if an image is captured from a flat surface (e.g., a screen) or not.

When working with light-field images it is possible to render 2D images that best describe the desired contents and then use traditional face analysis techniques. For instance, it is possible to use the refocusing functionality to create an all-in-focus 2D image, thus improving the analysis of faces that were captured at different focusing depths. Also, even if in principle depth and disparity express the same information, it can be useful to explore both types of information since different algorithms are used for their estimation. The alternative is to develop methods that directly explore the multidimensional light-field image, for instance, by processing the captured information as a tensor or as a multi-view array of 2D rendered images. The present paper overviews solutions of both types for a selected set of face analysis tasks.

The applications of face analysis are not limited to face recognition but also include other tasks such as face presentation attack detection (see Section 5), soft-biometrics categorization, and facial expression/emotion identification. The latter is the study of face features aiming to recognize the expression of the human face. Although this is a very studied field when dealing with traditional 2D cameras or other RGB-D or 3D sensors, until now, no works dealing with light field for human expression/emotion recognition have been published. Existing studies, employing other sensors, are mainly based on the extraction of face features and in the analysis of their changes when the face takes an expression. For an insight on recent techniques for facial expression recognition, the reader is referred to the survey paper by Corneanu et al. [19]. The present paper includes the review of a work dealing with face landmark localization on light-field images (see Section 3) that is an intermediary step for many face analysis applications, including face expression recognition.

Regarding soft-biometrics categorization, only one preliminary analysis work has been published so far. In [20], the authors observe a linear correlation between the depth of focus of the same face image, captured with a light-field camera and rendered at different focusing levels, and the gender and age scores obtained from it. Demonstrating the importance of controlling the image focus for this face analysis task.

As mentioned before, it is possible to exploit the rich information captured by light-field cameras to estimate depth and reconstruct 3D scenes. This feature, although not yet widely investigated, could be used to adapt and develop novel 3D face analysis techniques for light fields. A first step in this direction is described in [21], where the authors have compared the performances of a set of state-of-art algorithms for 2D and 3D face recognition on face images collected with two RGB-D sensors, which are a light-field camera and the Microsoft Kinect V1 (Read more at https://developer.microsoft.com/en-us/windows/kinect). Regarding the 3D information, the Kinect embeds a structured light 3D scanner. The 3D information is collected by projecting a known light pattern on the scene and analyzing how this pattern is deformed after hitting a surface. The results presented in [21], show that light fields are more robust than Kinect when dealing with facial occlusions (e.g., sunglasses). For a recent overview of 3D face recognition, the reader is referred to [22,23].

2.3. Databases

When discussing face analysis techniques, it is important to know which datasets were considered for their design, testing, and validation. By using publicly available databases it is possible to ensure that when face analysis solutions are reimplemented they achieve the initially reported performance results, while also guaranteeing the conditions for fair comparison among different techniques. Since this paper focuses on light-field face analysis, a summary of the existing light-field face databases is provided. These databases have been grouped into three categories:

- General-purpose light-field databases, which include facial images—These are databases initially developed with different purposes, but also happen to include face images. Metadata about the

faces is typically not available, but depending on the purpose of the algorithms being developed, these face images might be interesting to consider. A summary of the available databases is listed in Table 1.

- Light-field face databases—These are databases developed specifically to test face recognition solutions. Therefore, they typically include metadata information, such as the face bounding box and facial component coordinates, the subject gender, age and appearance (facial hair, makeup, haircut, earrings, necklace, scarf, piercings, scars, etc.). There are also databases that focus on specific facial components, such as the iris or the ear, which were derived from light-field face databases, and are also listed in Table 2.
- Light-field face presentation attack detection databases—These databases were specifically developed to test face presentation attack detection solutions. Several types of presentation attack instruments have been considered, such as printed paper, laptop, tablet, and smartphone. Images acquired with a light-field camera of the presentation attack instruments are then tested to check whether the light-field information is helpful in distinguishing a presentation attack from a genuine user presentation (*bona fide*). The available databases for this category are listed in Table 3.

Most of the listed databases are made available for research purposes. Other private datasets have been used in some works, such as [24], but they are not listed as they are not available for the research community. More detailed information about the content of each of the listed databases can be found in the provided references.

Table 1. General-purpose light-field databases that include facial images.

Name	Year	Image Acquisition	Spatial Resolution	Images with Faces	Content Variations
The (New) Stanford Light-Field Archive [25]	2008	Camera array	45 × 640 × 480	1 (students behind bushes)	Pose; distance; multiple people; occlusion; outdoor
EPFL Light-Field Image Dataset [26]	2016	Lytro Illum	15 × 15 × 625 × 434	18 (category: people)	Pose; distance; multiple people; outdoor
Stanford Lytro Light-Field Archive [27]	2016	Lytro Illum	15 × 15 × 625 × 434	17 (category: people)	Pose; distance; multiple people; occlusion; outdoor
SMART [28]	2016	Lytro Illum	15 × 15 × 625 × 434	1 (person)	Close-up, one person; reflection; indoor
Technicolor Light-Field dataset [29]	2017	4 × 4 camera rig	4 × 4 × 2048 × 1088	Several	Pose; distance; indoor

Table 2. Light-field face (and facial component) databases.

Name	Year	Image Acquisition	Image Type	Spatial Resolution	Test Subjects	Content Variations
GUCLF [30]	2013	2D camera; Lytro	2D; 2D rendered	5184 × 3456; 120 × 120	25	Pose; distance; illumination, multiple people; indoor; outdoor
LFC-MFD [31]	2016	2D camera; Lytro	2D; 2D rendered	5184 × 3456; 120 × 120	112	Pose; distance; illumination, multiple people; indoor; outdoor
IST-EURECOM LFFD [32]	2017	Lytro Illum	4D light field; 2D rendered; 2D depth map	15 × 15 × 625 × 434; 2022 × 1404; 2022 × 1404	100	Multiple sessions, pose, illumination, expression, occlusion
LFC-VID [31]	2016	Lytro	2D rendered	1080 × 1080	55 (iris)	Distance range: 9-15 inches; indoor
LLFEDB [33]	2018	Lytro Illum	4D light field; 2D rendered	15 × 15 × 625 × 434; 2022 × 1404	67 (ear)	Multiple sessions, pose, illumination, occlusion

Table 3. Light-field face presentation attack detection databases.

Name	Year	Image Acquisition	Image Type	Spatial Resolution	Test Subjects/Images	Attack Instruments
GUC-LiFFAD Database [34]	2015	Lytro	2D rendered (various depth/focus)	1080 × 1080	80 subjects/80 *bona fide* images; 2400 2D attack images	Printed paper: laser + inkjet, tablet
LLFFSD [35,36]	2018	Lytro Illum	4D light field; 2D rendered; 2D depth map	15 × 15 × 625 × 434; 2022 × 1404; 2022 × 1404	50 subjects/100 *bona fide* images; 600 attack images	Printed paper, wrapped printed paper, laptop, tablet, smartphone 1, smartphone 2
LLFEADB [37]	2018	Lytro Illum	4D light field; 2D rendered	15 × 15 × 625 × 434; 2022 × 1404	67 subjects/268 *bona fide* images; 1072 attack images	Laptop, tablet, smartphone 1, smartphone 2

3. Face Landmark Detection

Face landmark detection is the process of detecting salient points of the face, often corresponding to the edges of the main components of the face (e.g., mouth and eye corners, nose tip, chin, etc.). This is an intermediary step for many face-processing applications, including face recognition, face morphing, face liveness detection, and presentation attack detection. Recent research adopts face landmarks for face expression or emotion recognition by analyzing the dynamics of the landmarks while the face changes expression. Among more recreational applications, face landmarks are used on smartphones to superimpose, with surprising accuracy, funny masks or makeup or to apply "beauty filters" that operate differently on different areas of the face and thus require accurate face landmark detection.

Face landmark detection has proven extremely challenging due to the inherent face variability [38,39]. While achieving high accuracy on standard face images (e.g., frontal pose, neutral expression, standard illumination), state-of-the-art methods still suffer a significant drop in performance when dealing with different factors including pose, expression, illumination, and occlusions. An example of face landmark detection performed by different algorithms is presented in Figure 4.

To evaluate the performance of a face landmark detector, two different approaches can be defined: (i) compare the estimated landmarks with the ground truth, (ii) evaluate the performance for a specific task (e.g., emotion recognition). A straightforward way to assess landmark localization performance is to use manually annotated ground truths. A largely used metric is the normalized root mean square error (NRMSE). The normalization is typically done with respect to inter-ocular distance (IOD), which is defined as the distance between the two eye centers. Normalizing landmark localization errors by dividing by IOD makes the performance measure independent of the actual face size or the camera zoom factor [38]. The NRMSE between the ground-truth coordinates (x, y) and the estimated coordinates $(\tilde{x}, \tilde{y})$, is defined as:

$$\delta_i^k = \frac{d\{(x_i^k, y_i^k), (\tilde{x}_i^k, \tilde{y}_i^k)\}}{IOD} \tag{2}$$

where $d()$ indicates the Euclidean distance, k indicates the landmark index (e.g., eye corner, nose tip) and i is the image index. The overall landmark detector performance in terms of percentage of detected landmarks P, is computed by the following formula:

$$P = 100 \frac{\sum_{k=1}^{K} \sum_{i=1}^{N} [i : \delta_i^k < Th]}{K \times N} \tag{3}$$

where $[i : \delta_i^k < Th]$ is the indicator function of value 1 if the distance is smaller than Th, otherwise its value is 0. N denotes the number of test images and K the number of landmarks per face image. The threshold Th defines the error tolerance of the metric. The landmark estimation errors are assumed isotropic, so that one can conceive around each ground-truth landmark a detection circle with radius equal to the error threshold. The example in Figure 5 illustrates different error ranges corresponding to

different Th values. The detection threshold corresponds to a percentage of the IOD, and it is typically set to 10%, or below, of IOD (i.e., $Th <= 0.1$).

(**a**) (**b**) (**c**)

Figure 4. Face landmark detection: an example of output of three face landmark detectors. Face image extracted from the IST-EURECOM Light-Field Face Database (LFFD). (**a**) CLandmarks [40]; (**b**) IntraFace [41]; (**c**) DLIB.

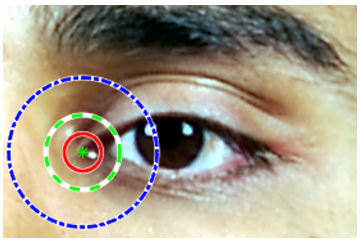


Figure 5. Detection thresholds of the eye inner corner. Concentric circles denote error ranges with radii 0.05 (red solid line), 0.1 (green dashed line), and 0.2 (blue dash-dotted line) times IOD, respectively. Eye image extracted from the IST-EURECOM LFFD.

The rich information provided by light-field data has pushed the researchers to consider how face landmark detection could be adapted to this new paradigm.

3.1. Facial-Landmark Localization Correction

The work presented in [42], explores the observation that in the EPI—described in Section 2.1—a 3D point is represented by a straight line. Thus, the position estimation of a face landmark on a light-field image can be optimized by forcing the estimated points to be on a straight line. Hereafter, such lines will be referred to as level lines, see Figure 6.

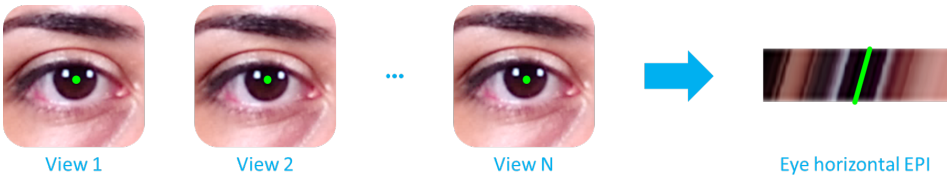


Figure 6. Level line: a 3D point lies on a straight line in the epipolar plane image. Eye image extracted from the IST-EURECOM LFFD.

The idea behind the work presented in [42] is thus to detect the EPI's level lines and use them to correct landmark position estimation to improve localization accuracy in terms of NRMSE.

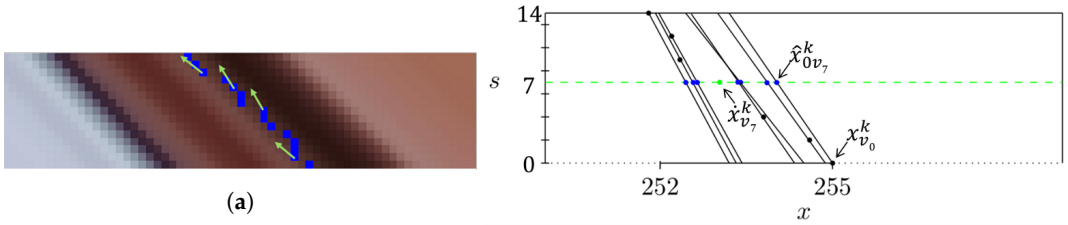
One way to detect the EPI's level lines, is to use structure tensors. The structure tensor, also referred to as the second-moment matrix, is derived from the gradient of the EPI as:

$$J_\sigma(x,u) = \Delta E_{y^*,v^*}(x,u) \cdot \Delta E_{y^*,v^*}(x,u)^T * G_\sigma \tag{4}$$

where G_σ is a Gaussian smoothing operator of variance σ^2. The orthogonal eigenvectors V_+ and V_- with respective eigenvalues λ_+ and λ_- (where $\lambda_+ > \lambda_-$) of $J(x,u)$ give a robust computation of the local gradient orientations locally at (x,u). The eigenvector V_- with the smallest eigenvalue describes the director vector of the level lines passing through (x,u).

3.1.1. Coordinates Correction

An ideal face landmark detector would localize the landmarks with perfect precision over all the views of the same light-field face image. This would create straight lines on the EPI, corresponding to the face landmarks. It is observed instead that even if the sub-aperture images are extremely similar—they only differ of a little horizontal and vertical disparity from each other—the landmark position estimation is different from view to view. This is even more evident on more challenging faces showing different expressions or pose. In the latter case, the estimated points would be scattered around the ground-truth level line. See the example given in Figure 7a.



(**a**) (**b**)

Figure 7. Coordinates correction: in (**a**), the scattered line produced by an estimated landmark point over the horizontal views; in (**b**), magnified view of the level lines computed using structure tensors.

The level lines corresponding to a landmark point k detected on the N views are computed using structure tensors and used to correct the coordinates of the landmark on the central view (i.e., view with index $N/2$).

An example of x coordinate correction is illustrated in Figure 7b. The same process is then applied to y coordinates. Please note that in the following formula and in Figure 7b the naming convention used in [42] is adopted: (s,t) denote the angular or view coordinates. The corrected coordinate $\dot{x}^k_{v_c}$ of the k^{th} landmark on the central view v_c, is obtained through a weighted sum of the projected points $\hat{x}^k_{s_{v_c}}$ that is the projection of the point $x^k_{v_s}$ from the view s, where $s = 0, 1, 2, ..., P$, on the central view (blue points in Figure 7b):

$$\dot{x}^k_{v_c} = \frac{\sum_{s=0}^{P} w^k_{v_s} \cdot \hat{x}^k_{s_{v_c}}}{\sum_{s=0}^{P} w^k_{v_s}}. \tag{5}$$

The weights are used to give more importance to points that are close to each other and less importance to eventual outliers. The weight $w^k_{v_s}$ is defined as the number of x-coordinates that are at a distance less than 0.1 pixel from the level line of the corresponding point $x^k_{v_s}$.

3.1.2. Results

Performance is assessed in terms of percentage of detected landmarks P on a set of 400 face images from the IST-EURECOM Light-Field Face Database (LFFD) database, including different face variations. Face landmarks are initially detected by DLIB (Read more about DLIB at http://blog.dlib.net/), a C++ library by Davis King that implements the method presented in [43] for face alignment based on regression trees. The results presented in Table 4 show that the method is particularly effective on more complex face variations, due to less accurate landmark localization of DLIB on these kinds of images. As a result, the correction is more beneficial.

Table 4. Face landmark correction results.

	Neutral Frontal Face		**Action Mouth Open**		**Pose Up Looking**		**Pose Half-Profile Left**	
	Original	Corrected	Original	Corrected	Original	Corrected	Original	Corrected
P(%)	97.81	**98.11**	95.70	**96.37**	91.80	**92.66**	77.68	**79.13**

4. Face Recognition

Automatic face recognition is becoming increasingly used for automated border control (ABC), access to protected areas or even to unlock smartphones. A wide literature on face recognition has been developed in recent decades, providing algorithms able to deal with challenging acquisition environments or with low-resolution images. With the improvement of optics, different sensors have been conceived, from time-of-flight cameras to light-field devices.

The works reviewed in this section adopt a different terminology when reporting performance. To provide a common understanding of the reported metrics, a standard terminology is used here, derived from the ISO/IEC 19795-1 [44]. Although accuracy is not a standard metric, we report it here as it is used in one of the reviewed works. Performance is reported in terms of the following metrics:

- **IR**: Identification Rate at rank 1–The (true-positive) identification rate at rank 1 is the proportion of identification transactions by a user enrolled in the system, for which user's true identifier is returned in first position in the candidate list. If the rank is omitted in the following, rank 1 is implied.
- **EER**: Equal Error Rate—value corresponding to $FMR = FNMR$
  - **FMR**: False Match Rate—proportion of the completed biometric non-mated comparison trials that result in a false match;
  - **FNMR**: False Non-Match Rate—proportion of the completed biometric mated comparison trials that result in a false non-match.
- **ACC**: Accuracy—corresponding to the average value of TMR and TNMR;
  - **TMR**: True Match Rate—proportion of the completed biometric mated comparison trials that result in a true match;
  - **TNMR**: True Non-Match Rate—proportion of the completed biometric non-mated comparison trials that result in a true non-match.

While most of the algorithms proposed for light-field face analysis are customized to be applied on data acquired with plenoptic cameras (Lytro devices most often), some works presented before 2006 use as database images collected with conventional sensors. In [1,2,4], Gross et al. suggest exploiting pose and illumination variations collected in the CMU PIE [45] and FERET databases [46] to create a light-field model of the face. Given a 2D image, the pixels belonging to the face are detected and the corresponding light-field image is computed as shown in Figure 8.

The same concept was explored in 2011 by Wibowo et al. [47] to perform face recognition from video sequences.

$$I - \sum_{i=1}^{d} \lambda_i W_i(\theta, \phi) = 0 \tag{6}$$

In [3], Zhou et al. integrate the Lambertian reflectance model to the method proposed by Gross et al. to consider illumination variations in addition to different poses.

The surveyed works in the following Subsections, employ light-field images collected with actual plenoptic cameras. They are organized in three categories depending on the image representation exploited. The categories are the following: Multi-focus-based methods; Sub-aperture-based methods, and deep-learning algorithms.

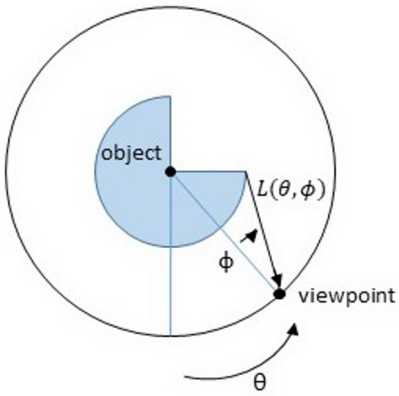


Figure 8. A visual description of the method used to define light field from 2D images in [1,2,4]. The object is conceptually placed within a circle. The angle to the viewpoint around the circle is measured by the angle θ, and the direction that the viewing ray makes with the radius of the circle is denoted ϕ. For each pair of angles θ and ϕ, the radiance of light reaching the viewpoint from the object is then denoted by $L(\theta;\phi)$, the light field.

4.1. Multi-Focus Based Methods

The first studies on face recognition on light-field images collected by plenoptic cameras, are based on the use of 2D images rendered from the light-field image at different focusing depths. In 2013, Raghavendra et al. [5] published an innovative technique to extract the best focused images from a set of 2D images from different depth planes. The authors presented one of the first light-field databases for face recognition [30] and proposed an approach to detect, select and extract features from light-field images (Figure 9). The main steps of the method are summarized below:

- The 2D images in the database are obtained by rendering the original light-field files at different focusing depths (Figure 10). All images are then processed with the Viola-Jones face detector [48] trained with 2429 face images and 3000 non-face samples. For each capture, the rendered image where the largest number of faces is detected, is chosen to define the facial regions.
- Once the faces are detected and cropped, the best image for each individual is selected according to an energy criterion. The authors chose as energy measure the 2D-Discrete Wavelet Transform (DWT) with Haar wavelet because of its robustness to noise and its content-independent property. The face image with larger energy is chosen.
- Local Binary Pattern (LBP) features [49] and Log-Gabor (LG) features [50] are extracted from the selected image. They are then used as input of Kernel Discriminant Analysis (KDA) [51] and Sparse Reconstruction Classifier (SRC) [52].

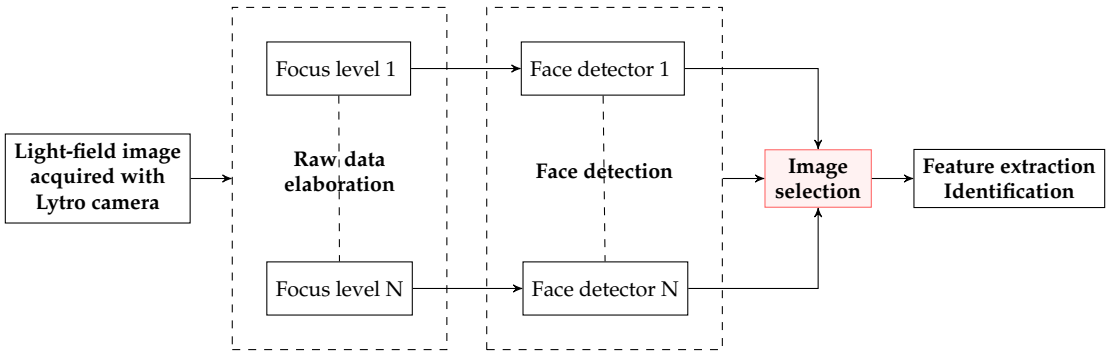


Figure 9. Workflow proposed in [5] .

Figure 10. An example of image rendered at different focus levels: The pictures are all extracted from the same light-field image. Image from the GUCLF database.

During the creation of the database, the images were acquired with both light-field and conventional cameras. The purpose of the work presented in [5] is to compare the identification rate obtained on light fields with the one achieved on standard 2D-images. The results show an increase in performance of 7.15% identification rate when using different state-of-the-art algorithm separately. The best-performing algorithm for 2D images is LG-SRC (53.97% IR), while for light field is LBP-SRC (61.12% IR). However, when the algorithms are fused at decision level, the improvement is only of 3.57%. Table 5 reports the obtained results.

In [53], the same authors investigate the use of light fields for identifying multiple faces present at different distance. From the GUCLF database, the depth information is exploited by combining the multi-focus images according to two strategies: to obtain an all-in-focus image; to obtain a super-resolution image. In this preliminary work, state-of-the-art super-resolution techniques [54–58] are used. The best performance is achieved using all-in-focus images and in the outdoor scenario with an identification rate at rank 1 of 53.62%. The best-performing super-resolution scheme, only achieves 32.86% identification rate. However, the authors do not compare the proposed methods with a 2D-image baseline.

A novel weighted image fusion scheme is proposed by Raghavendra et al. in [59]. While the face detection remains unchanged compared to the first work [5], image selection is based on entropy. For each focus version of each detected face, 2D-DWT is applied and the log-entropy is assessed with the following equation:

$$E = -\sum_{i=1}^{K} \left(log_2 \left(W_i \right)^2 \right) \tag{7}$$

where W_i for $i \in (1 : K)$ are the resulting wavelet coefficients.

Only samples with positive entropy are kept. The image entropy is normalized and sorted in decreasing order so that the best-in-focus image appears as first. The difference between adjacent entropy values (see Equation (8)) is used to assign a weight to each image.

$$D_j = |Sor_{j+1} - Sor_j| \quad for \quad j \in \{1 : \#images\} \tag{8}$$

$$w_j = \begin{cases} \left(0.5 + \left(0.5 * D_j\right)\right) * Max_w & if \quad D_j \geq Th \\ \frac{Max_w}{2} & otherwise \end{cases} \tag{9}$$

where Th is empirically set to 0.2 and Max_w is initially equal to 1 and then updated for each sample j with the weight value of the sample $j + 1$.

The image samples are then fused with the weighted sum rule (see Equation (10)).

$$W_f = \sum_j W_j * w_j \tag{10}$$

where W_j is the j^{th} image in the discrete wavelet domain and w_j the corresponding weight. The final result W_f is converted in spatial domain and used to extract features to perform face recognition.

The results obtained by using the proposed weighted image fusion scheme, are compared with the performance obtained by only selecting the image with largest entropy. The identification rate

achieved with the fusion scheme is higher in all the considered scenarios. The best result is achieved when LBP-SRC and LG-SRC are fused at decision level using the OR rule. The identification rate is of 75.12%. The proposed method is not compared with any 2D-image baseline.

The scheme is further improved in [60] where a new hybrid resolution enhancement technique is proposed. Also in this case, the first step of face detection remains unvaried. As in [59], 2D-DWT is performed on the images by applying filtering and downsampling, with high-pass (H) and low-pass (L) filters over rows and columns. This process produces four sub-images: I_{LL}, I_{LH}, I_{HL} and I_{HH}. For each sub-image, the wavelet energy is calculated and the image with largest energy is selected to represent the sample. The sub-image I_{LL}, containing the lower-frequency band, is replaced by a super-resoluted version of the original image obtained with a state-of-the-art method [54–58]. The obtained results show that the algorithm outperforms other well-known super-resolution techniques in terms of identification rate: 60.56% (Indoor), 58.87% (Corridor) and 51.47% (Outdoor).

In 2015, Raja et al. [61] studied the problem of light-field depth image fusion and in particular how to optimize the number of images to be fused. Following an approach inspired by [59], they consider in the fusion scheme only the two images with highest energy. As for [60], the energy is calculated from the energy sum of three sub-images obtained with low-pass and high-pass filtering. The best equal error rate of 4.14% is obtained with the combination of the proposed selection method and Laplacian Pyramid-based fusion using both sparse representation and multi-scale transform.

Table 5 reports the performance of the methods analyzed in this section.

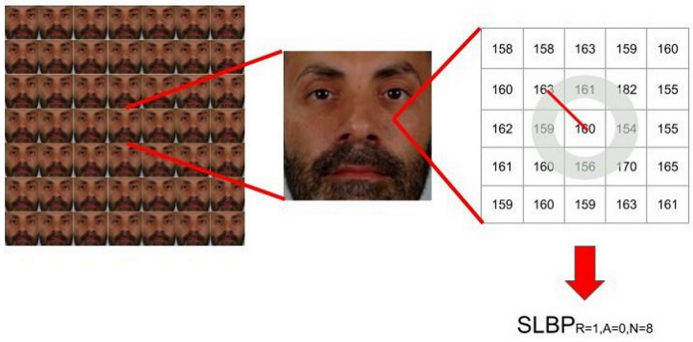
Table 5. Summary of multi-focus-based face recognition methods. Performance is assessed in terms of identification rate (IR) for all methods except for [61], where the equal error rate (EER) is reported. The abbreviations used in this table: WE–wavelet energy image selection; LBP–local binary pattern; LG–log-Gabor filter; LE–log-entropy image selection; SP–super-resolution; LE–log-entropy; LP–Laplacian pyramid image fusion; SRC–sparse reconstruction classifier; SRC–sparse reconstruction classifier; GUCLF–GUC light-field database.

Ref.	Year	Feature Extractor	Classifier	LF DB	2D Baseline	LF Perf.	Gain
[5]	2013	WE; LBP; LG	SRC	GUCLF	75.53% IR	79.10% IR	3.57%
[53]	2013	SR; LBP	SRC	GUCLF	-	53.62% IR	-
[59]	2013	LE; LBP; LG	SRC	GUCLF	-	75.12% IR	-
[60]	2013	LBP	SRC	GUCLF	-	60.56% IR	-
[61]	2015	LP; LBP	SRC	GUCLF	-	4.14% EER	-

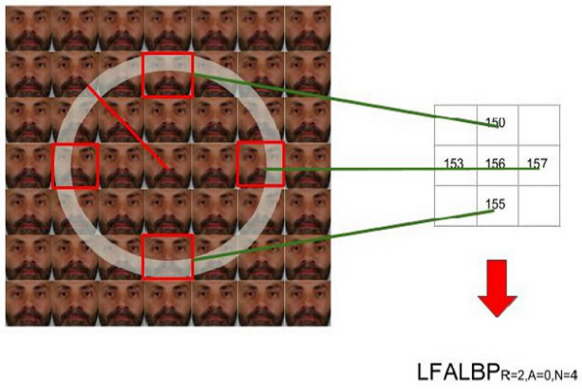
4.2. Sub-Aperture Based Methods

The publication of the IST-EURECOM LFFD [32] in 2017, has paved the way for the development of approaches based on the full information provided by light-field images. For example, the access to raw data allows investigation of the impact of sub-aperture representation on face recognition. So far, two main works have been proposed using the LFFD database, with the aim of improving face identification rate over standard algorithms based on 2D face images.

The first method is proposed by Sepas-Moghaddam et al. [62] and is inspired by LBP [49]. While for the computation of the classic LBP feature vector, adjacent pixels are considered, the Light-Field Local Binary Pattern (LFLBP) is composed by an additional component that includes the information stored in adjacent views (Figure 11):



(**a**)



(**b**)

Figure 11. Visual representation of (**a**) Spatial Local Binary Pattern (SLBP) and (**b**) Light-Field Angular Local Binary Pattern (LFALBP) used in [62].

- Spatial Local Binary Pattern (SLBP): the first component is the LBP feature vector extracted from the central view of the considered light-field image;
- Light-Field Angular Local Binary Pattern (LFALBP): the second component is a variation of classical LBP customized for light-field images. Let (x, y) be the reference sample and R the radius representing the distance of the selected adjacent views from the central view. Then, the LFALBP is defined as:

$$LFALBP_{R,A,N}(x,y) = \sum_{j=1}^{N} sign\left(L_{u,v,x,y} - L_{0,0,x,y}\right) * 2^{j-1} \tag{11}$$

where

$$\begin{cases} u = \left[Rsin\left(A + \frac{360^\circ}{N} * (j-1)\right)\right] \\ v = \left[Rcos\left(A + \frac{360^\circ}{N} * (j-1)\right)\right] \end{cases} \tag{12}$$

Angle A indicates the starting angle for the first sub-aperture image to consider in the angular neighborhood, and N indicates the number of views to consider. As in the conventional LBP descriptor, the binary thresholding result obtained by the *sign* function is multiplied by the binomial factor, 2^{j-1}, and the resulting values are summed to get the LFALBP pattern value for each sample position (x, y).

The authors have tested the proposed LFALBP with different sets of parameters and compared the proposed approach with several state-of-the-art techniques. The average results over three recognition tasks is of 92.1% identification rate, outperforming the best-performing 2D-image-based state-of-the-art method of 3%.

The second work is presented in [63]. As for [62], the authors exploit the sub-aperture representation of light-field face images. For each view of a light-field image, OpenFace (OF) [64], LBP [49] and LGBP [65] features are extracted. The impact of perspective shifting is demonstrated showing a linear relation between view shifting and face recognition dissimilarity score (see Figure 12).

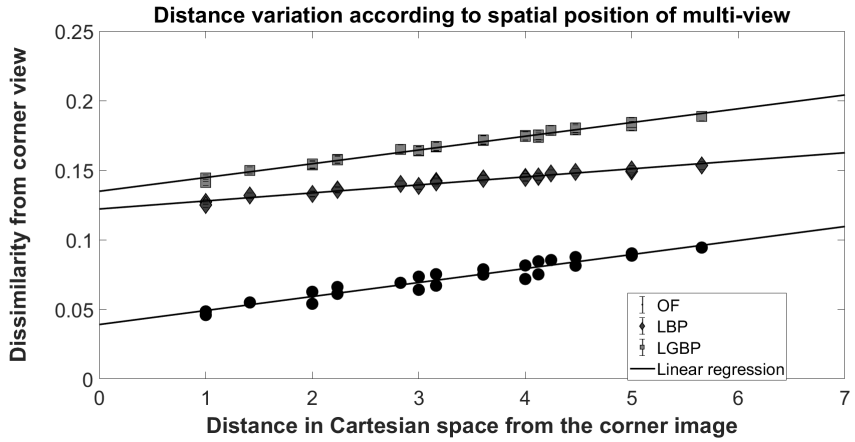


Figure 12. Relation between the distance of a pair of views in Cartesian space and the face recognition dissimilarity score: The observed linear correlation proves the feature extractor capability to capture the small face variation present in each view.

To evaluate the similarity of two light-field face images, all feature vectors extracted from the views of one image are matched with all feature vectors extracted from the other one. As comparison score, the Euclidean distance L^2 distance for OF features and chi-squared χ^2 distance for LBP and LGBP are used. Eight pseudo-distances are defined to combine the multiple values, obtained by comparing the features from the different views, to a single one.

Let A and B be feature vector sets describing the first and the second image, respectively, and A_C and B_C the feature vectors obtained from the corner view. The pseudo-distances are defined as:

- $d_{min} = min_{a \in A, b \in B} \left(d_s \left(a, b \right) \right)$
- $d_{minC} = min_{a \in A_C, b \in B_C} \left(d_s \left(a, b \right) \right)$
- $d_{mean} = mean \left(d_s \left(a, b \right) \right) \quad \forall a \in A, \forall b \in B)$
- $d_{meanC} = mean \left(d_s \left(a, b \right) \right) \quad \forall a \in A_C, \forall b \in B_C)$
- $d_{max} = max_{a \in A, b \in B} \left(d_s \left(a, b \right) \right)$
- $d_{maxC} = max_{a \in A_C, b \in B_C} \left(d_s \left(a, b \right) \right)$
- $d_{H_{mean}} = \frac{1}{\#A + \#B} \left(\sum_{a \in A} min_{b \in B} \left(d_s \left(a, b \right) \right) + \sum_{b \in B} min_{a \in A} \left(d_s \left(a, b \right) \right) \right)$
- $d_{H_{max}} = max \left(max_{a \in A} \left(min_{b \in B} \left(d_s \left(a, b \right) \right) \right), max_{b \in B} \left(min_{a \in A} \left(d_s \left(a, b \right) \right) \right) \right)$

The evaluation is conducted by comparing the results obtained by the proposed method with the results of standard 2D-image-based algorithms applied on the central view. Even though the baseline performances are already really good (EER lower than 0.05% when OF features are used), the authors prove that by using light field richer information it is possible to increase the performances by 0.53%.

Table 6 summarizes the performances of the methods described in this section.

Table 6. Summary of sub-aperture-based face recognition methods. The abbreviations used in this table: IR—identification rate; ACC—accuracy; LFLBP—light-field local binary pattern; OF—OpenFace; NN—nearest neighbor; D_{min}—minimum distance; LFFD—IST-EURECOM light-field face database.

Ref.	Year	Feature Extractor	Classifier	LF DB	2D Baseline	LF Perf.	Gain
[62]	2017	LFLBP	NN	LFFD	89.1% IR	92.1% IR	3%
[63]	2018	OF	D_{min}	LFFD	99.27% ACC	99.80% ACC	0.53%

4.3. Deep-Learning Algorithms

In 2018, Sepas-Moghaddam et al. used for the first time a deep-learning approach on light-field images to deal with the face recognition [66]. In this work, the authors fuse several representations of the raw data to exploit as much as possible the information stored in the image. Features are extracted by three VGG-Face neural networks [67] fused to feed a Support Vector Machine classifier (SVM). The main steps of the proposed approach are summarized below:

- The central view is the input of a pre-trained VGG-Face model;
- The disparity map calculated from the sub-aperture representation is used to fine-tune a second VGG-Face model;
- A third VGG-Face model is fine-tuned using depth maps.

Figure 13 illustrates the workflow of the method. The analysis of the rank-1 identification rate shows that the proposed algorithm, when fusing all available information, i.e., 2D+disparity+depth, achieves 98.1% identification rate. An improvement of 1.3% when compared to the use of 2D-image information only.

The same authors, in a second work, have developed a double-deep spatio-angular learning structure [68] based on the analysis of several images obtained from the sub-aperture representation. Each considered view is processed with a pre-trained VGG-Face network to extract a 4096-dimensional spatial feature vector. The output of this first elaboration is used as input for a Long Short-Term Memory (LSTM) Recurrent Neural Network [69] that extracts the angular dependencies across the views. The last step is a SoftMax classifier able to indicate the most probable identity of the individual represented in the image.

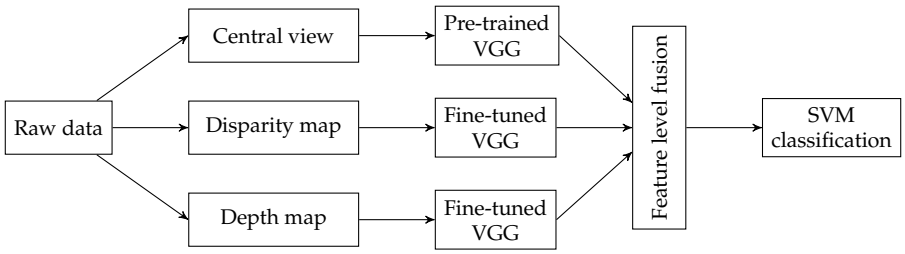


Figure 13. Workflow of the method proposed in [66].

The use of all views extracted from a light-field image would be computationally expensive and not necessary. The authors compare the application of this method on different set of views, selected based on different schemes. For example, the configuration called "Mid-density horizontal and vertical" is shown in Figure 14, where the selected views are those highlighted in red. The achieved identification rate is 98.60%, improving the performance of their previously proposed method of 1.2%, and of 5.7% when compared with the best-performing 2D-image-based approach.

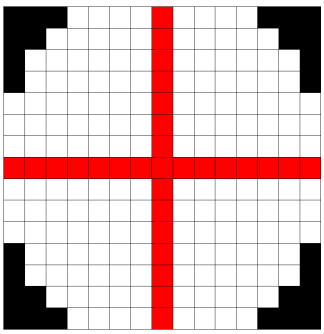


Figure 14. View selection scheme proposed in [68]: The red squares represent the position of the views selected in the mid-density horizontal and vertical configuration.

Table 7 reports the performances achieved by the methods described above.

Table 7. Summary of deep-learning-based face recognition methods. Abbreviation used in this table: IR–identification rate; VGG-Face–Visual Geometry Group's CNN descriptor; LSTM–long short-memory recurrent network; SVM–support vector machine; LFFD–IST-EURECOM Light-Field Face Database.

Ref.	Year	Feature Extractor	Classifier	LF DB	2D Baseline	LF Perf.	Gain
[66]	2018	VGG-Face	SVM	LFFD	96.8% IR	98.1% IR	1.3%
[68]	2018	VGG-Face + LSTM	SoftMax	LFFD	92.90% IR	98.60% IR	5.7%

5. Face Presentation Attack Detection

Besides the already mentioned airport application of ABC gates, access to mobile banking applications through face recognition, which by nature of the private interaction is unsupervised, motivates the importance of presentation attack detection. Over the years, academic and industry research centers developed countermeasures to detect biometric presentation attacks.

In general, a presentation attack can be conducted from an attacker with limited skills that interacts with a face capture device. The scenario of such attacks is well defined in the international standard that is entitled "ISO/IEC Information Technology–Biometric presentation attack detection" [70]. The intention of this standard is to provide a harmonized definition of terms and a taxonomy of attack techniques and a testing methodology that can evaluate PAD mechanisms.

5.1. Taxonomy for Presentation Attack Detection

In a multi-disciplinary community as in biometrics there is a tendency to struggle with a clear and non-contradictory use and understanding of its terms. Thus, ISO/IEC has undertaken significant efforts to develop a Harmonized Biometric Vocabulary (HBV) [71]. To formulate a common understanding of attacks on biometric systems the HBV was expanded with the following concepts that are provided in ISO/IEC 30107-1 Biometric presentation attack detection–Part1: Framework [70] and in ISO/IEC 30107-3 Biometric presentation attack detection–Part3: Testing and reporting [72]. Here, some of the standard terms are reported, which will be useful for the understanding of the following Sections:

- **presentation attack/attack presentation** : presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system
- **bona fide presentation**: interaction of the biometric capture subject and the biometric data capture subsystem in the fashion intended by the policy of the biometric system
- **presentation attack instrument (PAI)**: biometric characteristic or object used in a presentation attack
- **PAI species**: class of presentation attack instruments created using a common production method and based on different biometric characteristics
- **artefact**: artificial object or representation presenting a copy of biometric characteristics or synthetic biometric patterns
- **presentation attack detection (PAD)**: automated determination of a presentation attack

The framework defined in [70] considers two types of attacks. On the one hand, the *Active Impostor Presentation Attack* is considered, which attempts to subvert the correct and intended policy of the biometric capture subsystem and in which the attacker aims to be recognized as a specific subject known to the system (e.g., an impersonation attack). On the other hand, the *Identity Concealer Presentation Attack* as attempt of the attacker to avoid being matched to its own biometric reference in the system.

An attacker be it an active impostor or an identity concealer will use an object for their attack that is interacting with the capture device. Moreover, the potential of their attack will depend on their knowledge and the window of opportunity. However, for the object that is employed we can

anticipate for attacks against face capture devices as simple face image print outs or facial images being displayed on resolution tablets. Moreover, an attacker might present their genuine characteristic, but identification is avoided with non-conformant behavior with respect to the data capture regulations, e.g., by extreme facial expression or active visors.

5.2. Metrics for PAD Subsystem Evaluation

For a secure biometric system, the capture subsystem would be augmented with a PAD subsystem, which forwards the captured sample only then to the comparison subsystem, if it has been classified as bona fide. Such classification by the PAD subsystem is again subject to potential errors. Thus, when it comes to the testing of the detection subsystem ISO/IEC 30107-3 introduced metrics for PAD evaluation. A PAD subsystem shall be evaluated using two metrics namely [72]: (1) Attack Presentation Classification Error Rate (APCER): defined as the proportion of presentation attacks incorrectly classified as *Bona Fide* presentations (2) Bona Fide Presentation Classification Error Rate (BPCER): defined as the proportion of *Bona Fide* presentations incorrectly classified as presentation attacks.

The APCER can be calculated as follows:

$$APCER = \frac{1}{N_{PAIS}} \sum_{i=1}^{N_{PAIS}} (1 - RES_i) \tag{13}$$

where N_{PAIS} is the number of attack presentations for the given PAI [70]. RES_i takes the value 1 if the i^{th} presentation is classified as an attack presentation and value 0 if classified as *Bona Fide* presentation.

While the BPCER can be calculated as follows:

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}} \tag{14}$$

where N_{BF} is the number of *Bona Fide* presentations. RES_i takes the value 1 if the i^{th} presentation is classified as an attack presentation and value 0 if classified as *Bona Fide* presentation.

Some of the works presented in the following section report performance in terms of half total error rate of APCER and BPCER: $HTER = \frac{APCER+BPCER}{2}$. Please note that the HTER has been deprecated by the International Standard ISO/IEC 30107-3.

5.3. State-of-the-Art PAD with Light-Field Capture Devices

The vulnerability of face recognition capture devices for presentation attacks has intensively being discussed in the literature. For a recent overview, the reader is referred to [73], which also includes a discussion on countermeasures based on texture-based approaches or challenge response protocols. Given the low level of difficulty to render high-resolution video material on low-cost tablets, the limits to detect such attacks with a conventional 2D face capture device are obvious.

Soon after the light-field capture devices became available, they were investigated as means of defense against presentation attacks with 2D PAI. The motivation is straightforward as the light fields provide a superset of data acquired from the capture subject, which allows not only focus analysis at various depth and disparity exploitation.

Kim et al. [74] investigated two features to distinguish bona fide presentations from attack presentations conducted with printed PAD and high-resolution tablet. The first feature, namely the edge feature, is based on an extension of the LBP algorithm computed only on the areas located on the edge of the lower jaw. Whereas in [62] a similar LBP extension is proposed but the features are extracted from the sub-aperture representation, in [74] the inner and outer binary pattern methods are applied on microlens images. The microlens image is the raw information captured by the light-field camera sensor. It depicts the scene captured by each single microlens. This information is then processed to obtain the sub-aperture representation (see Figure 2). The authors define as inner binary pattern the LBP of average values evaluated for each microlens image. The outer binary pattern is computed

on surrounding microlens images. The second feature is based on sub-aperture representation, and is called ray-difference image. The ray-difference image is computed on the whole face area and is obtained by computing the difference between the central view and 4 other views (see Figure 15) at a given distance from the central one. The output are 4 error maps where the areas with larger depth change have the larger difference. The classical LBP algorithm is applied on the resulting images and the features are concatenated to feed an SVM. The work reported a half total error rate of APCER and BPCER that is in the range of 0.89% to 4.10% for the edge feature and 2.5% to 4.22% for the ray-difference feature.

In Raghavendra et al. [34] the focus variation between images at multiple depths is analyzed. As for [5], the authors render each light field at different focusing depths and detect faces in the obtained 2D images. Two schemes are proposed: the first explores the variation in focus among images extracted from the same raw data; the second is based on a simple decision rule on the number of depth images rendered by the Lytro camera on a single capture. The employed PAIs are collected in a dedicated dataset, namely the GUC Light-Field Face Artefact Database (GUC-LiFFAD), and include either high quality printed facial photos (printed using both laser and ink jet printers) and an electronic display. Detection accuracy is reported with the half total error rate of APCER and BPCER that varies from 4.01% to 5.27% depending on the PAI.

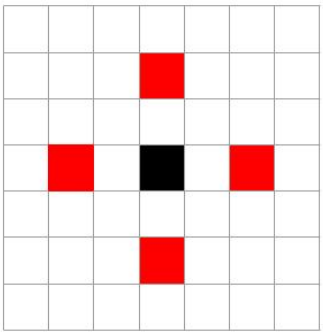


Figure 15. Map of views studied to evaluate the ray-difference features as described in [74].

Later, Ji et al. [75] proposed a PAD subsystem that is based on the Light-Field Histogram of Gradients (LFHOG) descriptor that extends the classical Histogram of Oriented Gradients (HOG) by considering gradients in three directions. While horizontal and vertical gradients are calculated as for standard HOG on the rendered 2D image, the gradient in depth direction is evaluated as the difference between two refocused images at different depths. Equation (15) describes the gradient in depth direction.

$$G_z(x,y) = I_{s_1}(x,y) - I_{s_2}(x,y) \tag{15}$$

where $I_s(x,y)$ is described in Equation (1). I_{s_1} and I_{s_2} are the refocused images in the reference depths s_1 and s_2, respectively. The module r and orientations θ and ϕ (in this case the orientations are two since the gradient vector is 3 dimensional) of the gradient are defined as in Equation (16).

$$\begin{aligned} r &= \sqrt{dx^2 + dy^2 + dz^2} \\ \theta &= \arccos \frac{dz}{r} \\ \phi &= \arctan \frac{dy}{dx} \end{aligned} \tag{16}$$

where (dx, dy, dz) are the components of the gradient of the light-field image in horizontal, vertical, and depth directions, respectively. The features are classified via SVM. The approach combines the analysis of the distribution of color intensity and of the distribution of scene depth. Detection accuracy is reported as 99.75% while omitting details on APCER and BPCER.

In 2018, Sepas-Moghaddam et al. [35,36] presented an overview of light-field-based PAD and developed novel methods exploiting the disparity information available in light-field images. The experiments are performed on the IST Lenslet Light-Field Face Spoofing Database (IST LLFFSD), the first face artefact database to include the raw light-field imaging files. In [35], they analyze color and texture variations associated with the different directions of light captured in light-field images. The algorithm is quite similar to the one for face recognition they previously presented in [62]: an extension of the classical LBP algorithm, customized for light-field images, is defined, namely the light-field angular LBP (LFALBP). Instead of considering adjacent pixels, values from adjacent views, transposed in the HVS and YC_bC_r color spaces, are used. In this scheme, two classifiers are created, one trained with $LFALBP_{HVS}$ features and one with $LFALBP_{YC_bC_r}$. Scores are merged to give the final classification result. Figure 16 shows a schematic representation of the framework. The proposed solution achieved the best performance compared to both 2D and LF-based state-of-the-art solutions. The average HTER detection error obtained ranges between 0.33% and 2.85%.

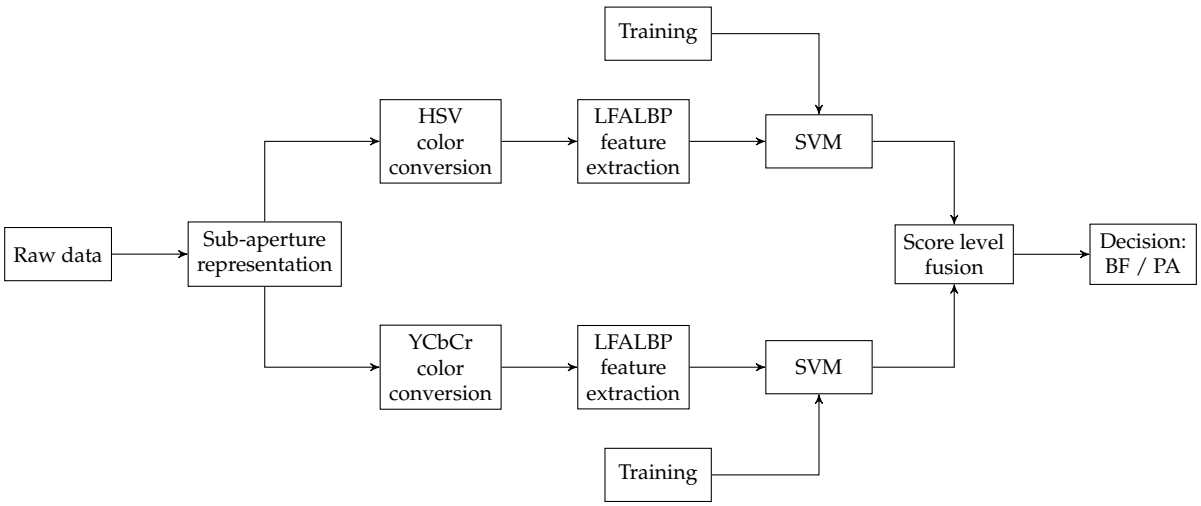


Figure 16. Schematic representation of the workflow for computing the $LBP_{HVS+YC_bC_r}$ features [35].

The method presented in [36] is inspired by the work in [35] but, instead of the LBP algorithm, it extends the well-known HOG method [76]. The authors propose the so-called Histogram of Disparity Gradients (HDG) descriptor, an extension of HOG targeting the description of light-field imaging disparity variations. Raw light-field data are pre-processed with the MATLAB Light-Field Toolbox: each image is represented as a 4-dimension matrix $L(u,v,x,y)$ where the first two values (u,v) indicate the considered sub-aperture and the last two values (x,y) the position of the pixel in the sub-aperture image. The horizontal $G_x(x,y)$ and vertical $G_y(x,y)$ disparity gradients are defined by Equation (17).

$$\begin{cases} G_x(x,y) = L(u_1,v_1,x,y) - L(u_2,v_2,x,y) \\ G_y(x,y) = L(u_3,v_3,x,y) - L(u_4,v_4,x,y) \end{cases} \tag{17}$$

The authors empirically prove that the most suitable sub-aperture images to perform PAD are: $(u_1 = 15, v_1 = 8)$, $(u_2 = 1, v_2 = 8)$, $(u_3 = 8, v_3 = 15)$, $(u_4 = 8, v_4 = 1)$ – where the sub-aperture image in the top-left corner has indexes $(1,1)$. Then, the disparity gradient magnitude $\bigtriangledown I$ and orientation θ are evaluated as in Equation (18).

$$\begin{cases} |\bigtriangledown I(x,y)| = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \\ \theta(x,y) = arctan(\frac{G_x(x,y)}{G_y(x,y)}) \end{cases} \tag{18}$$

The quantization, the normalization, and the concatenation are performed following the standard HOG algorithm. Compliant to the ISO/IEC standard [70] they report BPCER at a fixed 1% APCER. Detection accuracy for the set of presentation attack instruments (including laptop, tablet, mobile and paper) ranges from 0% to 0.45%.

In [77], a PAD method based on RGB-depth image pairs is presented. From each light-field raw file, a 2D-RGB image and a depth map of the captured face are computed. These images are already provided by the IST-EURECOM LFFD, but they can be easily computed from any light-field image. First, face landmark detection is performed on the 2D RGB image, identifying 68 landmarks for each face in the image. Then, depth values associated with each landmark are extracted from the corresponding depth map and used to define a feature vector called landmark depth feature (LDF). To evaluate the advantages of reducing the feature dimension, an alternative feature vector, namely Principal LDF (PLDF), composed by the first 10 principal components, is obtained by applying principal component analysis (PCA) on LDF. Figure 17 illustrates the workflow of the method. The two sets of features, LDF and PLDF, are tested separately. Three experiments show a HTER less than 1% for all the tested protocols, outperforming other PAD algorithms designed for light fields.

Convolutional neural networks (CNN) have been recently tested for PAD on light-field face images by Liu et al. in [24]. The proposed approach is based on two different features: the microlens image and the ray-difference image. The ray-difference image is computed, as described above, using the views at distance 3 from the central view. The 4 error maps obtained are concatenated to be fed into the CNN. The CNN is configured to process an input of size $250 \times 250 \times 3$, where 3 corresponds to the RGB color channels. The CNN is implemented on the tensorflow platform. Experiments are carried out on a dataset collected by the authors using a Lytro ILLUM camera and including the following PAIs: printed photo, warped photo, and screen-displayed photo. The results are reported in terms of HTER, and with the microlens feature they reach 0.028%.

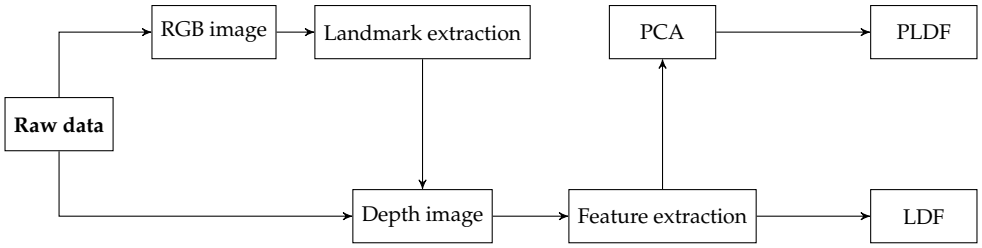


Figure 17. Workflow of the PAD method proposed in [77] .

Table 8 summarizes the performances of the methods described in this section.

What is lacking for light-field PAD research is a comparative benchmark on a database that includes sophisticated silicone mask [78], which should be considered now state of the art in face presentation attacks.

Table 8. Summary of presentation attack detection methods. Abbreviations used in this table: EF—edge feature; RD—ray-difference image; FV—focus variation; DR—decision rule; LFHOG—light-field histogram of oriented gradients; LFALBP—light-field angular local binary pattern; HDG—histogram of disparity gradients; LDF—landmark depth feature; PLDF—principal landmark depth feature; CNN—convolutional neural networks; SVM—support vector machine; GUC-LiFFAD—GUC Light-Field Face Artefact Database; IST LLFFSD—IST Lenslet Light-Field Face Spoofing Database.

Ref.	Year	Feature Extractor	Classifier	LF DB	LF Perf.
[74]	2014	EF + RD	SVM	Private	0.89–4.22% HTER
[34]	2015	FV + DR	SVM	GUC-LiFFAD	4.01–5.27% HTER
[75]	2016	LFHOG	SVM	Private	99.75% ACC
[35]	2018	$LFALBP_{HVS}$ + $LFALBP_{YC_bC_r}$	SVM	IST LLFFSD	0.33–2.85% HTER
[36]	2018	HDG	SVM	IST LLFFSD	0–0.45% BPCER @ 1%APCER
[77]	2018	LDF; PLDF	SVM	IST LLFFSD	0–0.8% HTER
[24]	2019	CNN	SVM	Private	0.028% HTER

6. Conclusions

As soon as new visual sensors are placed on the market, it is usual that several R&D teams revisit existing methods to process the new image formats in terms of transmission, compression, and displaying. It was the case, some years ago, for Microsoft Kinect imaging sensors [79]. The Joint Photographic Experts Group (JPEG) Standardization committee has defined the JPEG Pleno to provide a standard framework for representing new imaging modalities, such as texture-plus-depth, point cloud, holographic imaging, and light fields.

Apart from the need for purely adapting existing processing methods to the new image format, it is also of interest to investigate to what extent such new visual sensors, which carry richer information and provide additional features, can increase the performance of some existing tasks related to computer vision, e.g., facial image processing. Recently, some companies (e.g., Lytro, Raytrix) have developed imaging sensors integrating light-field technology, the so-called plenoptic camera. This camera can capture in one shot multiple views of the scene, offering new capabilities in terms of depth estimation and image refocusing.

The present article reviews methods that deal with face image processing including landmark detection, recognition, and presentation attack detection in the context of acquisitions performed by Lytro cameras (almost all existing databases have been collected using Lytro products). It has been proven that the richer information provided by such new sensors offers new capabilities in several existing domains. Regarding face recognition, the use of light-field imaging can contribute to increasing performance but with a non-negligible additional computation complexity. As showed in Tables 5–7, the performance gain ranges from 0.53% to 5.7%. At the moment, it is hard to predict the future of such new devices in real applications for face recognition.

The results obtained from the application of light fields for face presentation attack detection are encouraging. The provided depth information makes light-field-based PAD solutions robust to 2D presentation attacks instruments. This is of particular importance in unsupervised authentication scenarios, where the attacker can easily present a 2D replica of an authorized user face without being detected. However, as previously mentioned in Section 5, it would be of great interest to investigate more sophisticated PAIs, such as silicone masks. If we consider other fields of research, light fields have proven to be useful for accurate millimetric measurement for automated visual inspection. In [80] the use of refocused depth calibrated images allows performance of measurements on industrial objects and the authors also show that a pixel metric scale can be estimated at different depths, avoiding the use of other measurement devices. It would be interesting to see if these techniques could be successfully applied for PAD in presence of 3D face masks.

Although light fields have not been applied in this field yet, the authors believe that a promising employment of such technology is for facial expression or micro-expression analysis. Section 3 has illustrated the possible benefits of using light-field images for more accurate face landmark detection. The use of light-field videos would also allow the study of the dynamic of the landmarks while the face changes expression.

Light-field devices able to record videos have already been deployed by Raytrix (Read more at https://raytrix.de/). It is safe to assume that in the future more devices will be developed and that they will integrate more features, such as near infrared and thermal imaging sensors. The computational complexity of light-field pictures comes also from the number of microlenses in the microlens array mounted on the plenoptic camera. Some tasks may require the use of only a small number of views. Some preliminary results obtained in [80,81] lead to reduce the number of views. A custom prototype of light-field camera is obtained by putting a small array of microlenses (e.g., of size 2×2) in front of the imaging sensor. In this case, 3D information is sparse but preserved. Concerning Lytro, to date the American company has ceased operations in late March 2018.

Finally, as demonstrated by this survey, light-field face analysis so far has been assessed almost only in comparison to 2D face analysis. A future line of study could therefore investigate the benefits of light fields for face analysis when compared to other imaging technologies, including but not limited to texture-plus-depth, point cloud, thermal, and high dynamic range.

Author Contributions: Author Contributions: conceptualization, V.C., C.GA. and J.D.; abstract and introduction, C.GA.; overview on "Light Fields", C.GU., V.C., and C.GA.; overview on "Face Analysis " and “Databases”, P.L.C. and C.GA.; survey on "Face landmark detection", C.GA.; survey on "Face recognition", V.C.; survey on "Face presentation attack detection", C.B. and V.C.; conclusions, C.GA. and J.D.; writing–original draft preparation, C.GA., V.C., C.GU., P.L.C., and C.B.; writing–review and editing, C.GA.; supervision, P.L.C., C.B., J.D., and C.GU.

Funding: This research was funded by EU H2020 Research and Innovation Programme under grant agreement no. 700259 (PROTECT) and EU H2020 Research and Innovation Programme under grant agreement no. 694122 (ERC advanced grant CLIM).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gross, R.; Matthews, I.; Baker, S. Eigen light-fields and face recognition across pose. In Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, 21–21 May 2002; pp. 3–9.
2. Gross, R.; Matthews, I.; Baker, S. Fisher Light-Fields for Face Recognition Across Pose and Illumination. In Proceedings of the German Symposium on Pattern Recognition (DAGM), Zurich, Switzerland, 16–18 September 2002.
3. Zhou, S.; Chellappa, R. Illuminating light field: Image-based face recognition across illuminations and poses. In Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 19 May 2004; pp. 229–234.
4. Gross, R.; Matthews, I.; Baker, S. Appearance-based face recognition and light-fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 449–465, doi:10.1109/TPAMI.2004.1265861.
5. Raghavendra, R.; Yang, B.; Raja, K.B.; Busch, C. A new perspective—Face recognition with light-field camera. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–8. doi:10.1109/ICB.2013.6612980.
6. Levoy, M.; Hanrahan, P. Light Field Rendering. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; ACM: New York, NY, USA, 1996; pp. 31–42.
7. Gortler, S.; Grzeszczuk, R.; Szeliski, R.; Cohen, M. The Lumigraph. In Proceedings of the SIGGRAPH, New Orleans, LA, USA, 4–9 August 1996; pp. 43–54.
8. Wilburn, B.; Joshi, N.; Vaish, V.; Talvala, E.V.; Antunez, E.; a. Barth.; Adams, A.; Horowitz, M.; Levoy, M. High Performance Imaging Using Large Camera Arrays. *ACM Trans. Gr.* **2005**, *24*, 765–776.

9. Unger, J.; Gustavson, S.; Larsson, P.; Ynnerman, A. Free Form Incident Light Fields. *Comput. Gr. Forum* **2008**, doi:10.1111/j.1467-8659.2008.01268.x.
10. Adelson, T.; Wang, J. Single Lens Stereo with a Plenoptic Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 99–106.
11. Ng, R.; Levoy, M.; Bredif, M.; Duval, G.; Horowitz, M.; Hanrahan, P. Light field photography with a hand-held plenoptic camera. In *Stanford Tech Report CTSR 2005-02*; 2005; Volume 2, pp. 1–11.
12. Lumsdaine, A.; Georgiev, T. The focused plenoptic camera. In Proceedings of the IEEE International Conference on Computational Photography, San Francisco, CA, USA, 16–17 April 2009; pp. 1–8.
13. Bolles, R.C.; Baker, H.H.; Marimont, D.H. Epipolar-plane image analysis: An approach to determining structure from motion. *Int. J. Comput. Vis.* **1987**, *1*, 7–55, doi:10.1007/BF00128525.
14. Jeon, H.G.; Park, J.; Choe, G.; Park, J.; Bok, Y.; Tai, Y.W.; So Kweon, I. Accurate depth map estimation from a lenslet light field camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1547–1555.
15. Huang, C.T. Empirical Bayesian Light-Field Stereo Matching by Robust Pseudo Random Field Modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 552–565.
16. Jiang, X.; Le Pendu, M.; Guillemot, C. Depth estimation with occlusion handling from a sparse set of light field views. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 634–638.
17. Wanner, S.; Goldluecke, B. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans Pattern Anal. Mach. Intell.* **2013**, *36*, 606–619.
18. Zhang, S.; Sheng, H.; Li, C.; Zhang, J.; Xiong, Z. Robust depth estimation for light field via spinning parallelogram operator. *J. Comput. Vis. Image Underst.* **2016**, *145*, 148–159.
19. Corneanu, C.A.; Simón, M.O.; Cohn, J.F.; Guerrero, S.E. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1548–1568, doi:10.1109/TPAMI.2016.2515606.
20. Chiesa, V.; Dugelay, J. Impact of multi-focused images on recognition of soft biometric traits. In *Applications of Digital Image Processing XXXIX*; International Society for Optics and Photonics, Bellingham, Washington USA; SPIE: 2016; Volume 9971, p. 99710Q.
21. Chiesa, V.; Dugelay, J. Kinect vs Lytro in RGB-D Face Recognition. In Proceedings of the 2018 International Conference on Cyberworlds (CW), Singapore, 3–5 October 2018; pp. 345–350. doi:10.1109/CW.2018.00069.
22. Soltanpour, S.; Boufama, B.; Jonathan Wu, Q. A survey of local feature methods for 3D face recognition. *Pattern Recognit.* **2017**, *72*, 391–406, doi:10.1016/j.patcog.2017.08.003.
23. Ouyang, S.; Hospedales, T.; Song, Y.Z.; Li, X.; Loy, C.; Wang, X. A survey on heterogeneous face recognition: Sketch, infra-red, 3D and low-resolution. *Image Vis. Comput.* **2016**, *56*, 28–48, doi:10.1016/j.imavis.2016.09.001.
24. Liu, M.; Fo, H.; Wei, Y.; Rehman, Y.; Po, L.; Lo, W. Light field-based face liveness detection with convolutional neural networks. *SPIE Electron. Imaging* **2019**, *28*, 013003.
25. Vaish, V.; Adams, A. The (New) Stanford Light Field Archive. 2008. Available online: http://lightfield.stanford.edu/lfs.html (accessed on 20 February 2019).
26. Řeřábek, M.; Ebrahimi, T. New Light Field Image Dataset. In Proceedings of the 8th International Workshop on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.
27. Raj, A.S.; Lowney, M.; Shah, R.; Wetzstein, G. The Stanford Lytro Light Field Archive. 2016. Available online: http://lightfields.stanford.edu/LF2016.html (accessed on 20 February 2019).
28. Paudyal, P.; Olsson, R.; Sjöström, M.; Battisti, F.; Carli, M. SMART: A light field image quality dataset. In Proceedings of the 7th International Conference on Multimedia Systems (ICMS), Klagenfurt, Austria, 10–13 May 2016.

29. Sabater, N.; Boisson, G.; Vandame, B.; Kerbiriou, P.; Babon, F.; Hog, M.; Langlois, T.; Gendrot, R.; Bureller, O.; Schubert, A.; Allie, V. Dataset and Pipeline for Multi-View Light-Field Video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.
30. Raghavendra, R.; Raja, K.B.; Yang, B.; Busch, C. GUCLF: A new light field face database. In Proceedings of the 8th Iberoamerican Optics Meeting and 11th Latin American Meeting on Optics, Lasers, and Applications, Porto, Portugal, 22–26 July 2013; Volume 8785, p. 8785. doi:10.1117/12.2026184.
31. Raghavendra, R.; Raja, K.; Busch, C. Exploring the usefulness of light field cameras for Biometrics: An Empirical Study on Face and Iris Recognition. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 922–936.
32. Sepas-Moghaddam, A.; Chiesa, V.; Correia, P.L.; Pereira, F.; Dugelay, J. The IST-EURECOM Light Field Face Database. In Proceedings of the 2017 5th International Workshop on Biometrics and Forensics (IWBF), Coventry, UK, 4–5 April 2017; pp. 1–6. doi:10.1109/IWBF.2017.7935086.
33. Sepas-Moghaddam, A.; Pereira, F.; Correia, P. Ear recognition in a light field imaging framework: A new perspective. *IET Biom.* **2018**, *7*, 224–231.
34. Raghavendra, R.; Raja, K.; Busch, C. Presentation Attack Detection for Face Recognition Using Light Field Camera. *IEEE Trans. Image Process.* **2015**, *24*, 1060–1074.
35. Sepas-Moghaddam, A.; Malhadas, L.; Correia, P.; Pereira, F. Face Spoofing Detection using a Light Field Imaging Framework. *IET Biom.* **2018**, *7*, 39–48.
36. Sepas-Moghaddam, A.; Pereira, F.; Correia, P. Light Field based Face Presentation Attack Detection: Reviewing, Benchmarking and One Step Further. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1696–1709.
37. Sepas-Moghaddam, A.; Pereira, F.; Correia, P. Ear Presentation Attack Detection: Benchmarking Study with First Lenslet Light Field Database. In Proceedings of the 26th European Signal Processing Conference (EUSIPCO 2018), Rome, Italy, 3–7 September 2018.
38. Çeliktutan, O.; Ulukaya, S.; Sankur, B. A comparative study of face landmarking techniques. *EURASIP J. Image Video Process.* **2013**, *2013*, 13, doi:10.1186/1687-5281-2013-13.
39. Johnston, B.; Chazal, P.d. A review of image-based automatic facial landmark identification techniques. *EURASIP J. Image Video Process.* **2018**, *2018*, 86, doi:10.1186/s13640-018-0324-4.
40. Uřičář, M.; Franc, V.; Thomas, D.; Sugimoto, A.; Hlaváč, V. Multi-view facial landmark detector learned by the Structured Output SVM. *Image Vis. Comput.* **2016**, *47*, 45–59. doi:10.1016/j.imavis.2016.02.004.
41. Xiong, X.; De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013.
42. Galdi, C.; Younes, L.; Guillemot, C.; Dugelay, J.L. A new framework for optimal facial landmark localization on light-field images. In Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2018; pp. 1–4.
43. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874, doi:10.1109/CVPR.2014.241.
44. ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2006. Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework*. International Organization for Standardization and International Electrotechnical Committee, Geneva, Switzerland: 2006.
45. Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.A. Multi-PIE. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008.
46. Rauss, P.J.; Phillips, J.; Moon, H.; Rizvi, S.; Hamilton, M.K.; Trent DePersia, A. The FERET (face recognition technology) program. In Proceedings of the 25th Annual AIPR Workshop on Emerging Applications of Computer Vision, Washington, DC, USA, 1996; doi:10.1117/12.267831.
47. Wibowo, M.E.; Tjondronegoro, D. Face Recognition across Pose on Video Using Eigen Light-Fields. In Proceedings of the 2011 International Conference on Digital Image Computing: Techniques and Applications, Noosa, QLD, Australia, 6–8 December 2011; pp. 536–541, doi:10.1109/DICTA.2011.96.
48. Viola, P.; Jones, M. Robust Real-Time Face Detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154, doi:10.1109/ICCV.2001.937709.

49. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987, doi:10.1109/TPAMI.2002.1017623.
50. Field, D.J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **1987**, *4*, 2379–2394, doi:10.1364/JOSAA.4.002379.
51. Lu, J.; Plataniotis, K.N.; Venetsanopoulos, A.N. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. Neural Netw.* **2003**, *14*, 117–126, doi:10.1109/TNN.2002.806629.
52. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227, doi:10.1109/TPAMI.2008.79.
53. Raghavendra, R.; Raja, K.B.; Yang, B.; Busch, C. Improved face recognition at a distance using light field camera and super resolution schemes. In Proceedings of the 6th International Conference on Security of Information and Networks, Aksaray, Turkey, 26–28 November 2013 .
54. Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP Graph. Model. Image Process.* **1991**, *53*, 231–239, doi:10.1016/1049-9652(91)90045-L.
55. Stark, H.; Oskoui, P. High-resolution image recovery from image-plane arrays, using convex projections. *J. Opt. Soc. Am. A Opt. Image Sci.* **1989**, *6*, 1715–26, doi:10.1364/JOSAA.6.001715.
56. W. Gerchberg, R. Super-resolution through Error Energy Reduction. *Opt. Acta Int. J. Opt.* **1974**, *21*, 709–720, doi:10.1080/713818946.
57. Papoulis, A. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Trans. Circuits Syst.* **1975**, *22*, 735–742, doi:10.1109/TCS.1975.1084118.
58. Zomet, A.; Rav-Acha, A.; Peleg, S. Robust super-resolution. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii USA, 8–14 December 2001; Volume 1, doi:10.1109/CVPR.2001.990535.
59. Raghavendra, R.; Raja, K.B.; Yang, B.; Busch, C. A novel image fusion scheme for robust multiple face recognition with light-field camera. In Proceedings of the 16th International Conference on Information Fusio, Istanbul, Turkey, 9–12 July 2013; pp. 722–729.
60. Raghavendra, R.; Raja, K.B.; Yang, B.; Busch, C. Comparative evaluation of super-resolution techniques for multi-face recognition using light-field camera. In Proceedings of the 2013 18th International Conference on Digital Signal Processing (DSP), Santorini, Greece, 1–3 July 2013; pp. 1–6, doi:10.1109/ICDSP.2013.6622829.
61. Raja, K.B.; Raghavendra, R.; Alaya Cheikh, F.; Busch, C. Evaluation of fusion approaches for face recognition using light field cameras. In Proceedings of the 2015 Colour and Visual Computing Symposium (CVCS), Gjovik, Norway, 25–26 August 2015, doi:10.1109/CVCS.2015.7274896.
62. Sepas-Moghaddam, A.; Correia, P.L.; Pereira, F. Light field local binary patterns description for face recognition. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3815–3819. doi:10.1109/ICIP.2017.8296996.
63. Chiesa, V.; Dugelay, J.L. On Multi-View Face Recognition Using Lytro Images. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2250–2254, doi:10.23919/EUSIPCO.2018.8553572.
64. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications*; Technical report, CMU-CS-16-118; CMU School of Computer Science: Pittsburgh, PA, USA, 2016.
65. Zhang, W.; Shan, S.; Gao, W.; Chen, X.; Zhang, H. Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; Volume 1, pp. 786–791, doi:10.1109/ICCV.2005.147.
66. Sepas-Moghaddam, A.; Correia, P.; Nasrollahi, K.; Moeslund, T.; Pereira, F. In Proceedings of the Light Field Based Face Recognition via a Fused Deep Representation, Aalborg, Denmark, 17–20 September 2018; pp. 1–6, doi:10.1109/MLSP.2018.8516966.
67. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
68. Sepas-Moghaddam, A.; Haque, M.A.; Correia, P.L.; Nasrollahi, K.; Moeslund, T.B.; Pereira, F. A Double-Deep Spatio-Angular Learning Framework for Light Field based Face Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, doi:10.1109/TCSVT.2019.2916669.

69. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–80, doi:10.1162/neco.1997.9.8.1735.
70. ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Information Technology–Biometric Presentation Attack Detection—Part 1: Framework*. International Organization for Standardization, Geneva, Switzerland: 2016.
71. ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 2382-37:2017 Information Technology–Vocabulary—Part 37: Biometrics*. International Organization for Standardization, Geneva, Switzerland: 2017.
72. ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-3. Information Technology–Biometric Presentation Attack Detection—Part 3: Testing and Reporting*. International Organization for Standardization, Geneva, Switzerland: 2017.
73. Ramachandra, R.; Busch, C. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Comput. Surv.* **2017**, *50*, 8:1–8:37, doi:10.1145/3038924.
74. Kim, S.; Ban, Y.; Lee, S. Face liveness detection using a light field camera. *Sensors* **2014**, *14*, 22471–22499, doi:10.3390/s141222471.
75. Ji, Z.; Zhu, H.; Wang, Q. LFHOG: A discriminative descriptor for live face detection from light field image. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1474–1478, doi:10.1109/ICIP.2016.7532603.
76. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
77. Chiesa, V.; Dugelay, J.L. Advanced face presentation attack detection on light field images. In Proceedings of the 17th International Conference of the Biometrics Special Interest Group, BIOSIG, Darmstadt, Germany, 26–28 September 2018.
78. Bhattacharjee, S.; Mohammadi, A.; Marcel, S. Spoofing Deep Face Recognition with Custom Silicone Masks. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–8.
79. Special issue on computer vision for RGB-D sensors: Kinect and its applications. *IEEE Trans. Cybern.* **2012**, *42*, 1295–1296, doi:10.1109/TSMCB.2012.2207010.
80. Riou, C.; Colicchio, B.; Lauffenburger, J.P.; Cudel, C. Interests of refocused images calibrated in depth with a multi-view camera for control by vision. In *Thirteenth International Conference on Quality Control by Artificial Vision 2017*; International Society for Optics and Photonics, Bellingham, Washington USA: 2017; Volume 10338, p. 1033807.
81. Gendre, L.; Bazeille, S.; Bigué, L.; Cudel, C. Interest of polarimetric refocused images calibrated in depth for control by vision. In *Unconventional Optical Imaging*; International Society for Optics and Photonics, Bellingham, Washington USA: 2018; Volume 10677, p. 106771W.